

*Library Review*, 2001, Vol. 50, No.1, p16-27.

ISSN: 0024-2535

DOI: 10.1108/00242530110363217

<http://www.emeraldinsight.com/>

<http://www.emeraldinsight.com/journals.htm?issn=0024-2535&volume=50&issue=1>

© 2001 MCB UP Ltd

## Metadata and the future of cataloging

**Magda El-Sherbini**

### Abstract

This article is a survey of representative metadata efforts comparing them to MARC 21 metadata in order to determine if new electronic formats require the development of a new set of standards. This study surveys the ongoing metadata projects in order to identify what types of metadata exist and how they are used and also compares and analyzes selected metadata elements in an attempt to illustrate how they are related to MARC 21 metadata format elements.

### Introduction and definition of metadata

Metadata has been the centre of attention inside and outside the library community for several years. It has been the subject of many discussions and poster sessions at the American Library Association meetings (for example The Committee on Cataloging: Description and Access (CC:DA) and Machine Readable Bibliographic Information (MARBI)). Several workshops have been conducted to give an overview of metadata and to define the term and its use. So what is metadata?

*A User Guide for Simple Dublin Core* provides the following definition:

Metadata describes an information resource. The term "Meta" comes from a Greek word that denotes something of a higher or more fundamental nature. Metadata, then, is data about data. It is the Internet-age term for information that librarians traditionally have put into catalogs and it most commonly refers to descriptive information about Web resources [1].

In "Computer Wire: Data Warehousing Tools Bulletin: Briefing Paper: What is Metadata" [2] the authors introduce and define metadata from the perspective of the requirements and demands of data warehousing. They explain why metadata is an essential part of a practical data warehouse and give examples of how metadata can help to make the warehouse user-friendly. They describe metadata as an abstraction from data. "It is high-level data that describes lower-level data. The card catalog is a metadata that describe the collection of a library,"

In the definition provided by Milstead and Feldman the emphasis is placed on the variety of digital formats that can be defined by metadata:

The term is generally applied to electronic resources (though it does not have to be) and refers to "data" in the broadest sense; data-sets, textual information, graphics, music, and anything else that is likely to appear electronically. While the concept includes indexing and cataloging information (information for "resource discovery" in Webspeak), it can go far beyond conventional document representations, such as MARC records [3].

This small sampling of approaches to the definition of metadata provides a glimpse of the spectrum of issues and factors that contribute to the phenomenon. Most authors agree that metadata is a term that is used to denote information about electronic information resources. It is now generally acknowledged as being vital in order to enable effective retrieval of information in our increasingly complex, digitally-networked information environment.

The most common definition of the term "metadata" is data about data - information that describes other information. If, for example, a Web page has an author, a title, a date of creation, and a unique Internet address, these elements constitute metadata about this page.

Although the term is relatively new to library terminology, the concept of describing data or information has been with us since the very first efforts at classifying and organizing information. In fact all the library tools, which have been used to describe contents of data, can be called metadata. AACR2, for example, is a tool designed to help organize access to data. It is a code of rules for creating bibliographic descriptions of documents, a classification system created to simply provide access to materials.

This article surveys representative metadata efforts and compares them to MARC 21 metadata in order to determine if new electronic formats such as Web pages, electronic journals, etc., require the development of a new set of standards. The article addresses the question of the originality of several metadata standards and their general usefulness. The study employs the following methods:

- survey of ongoing metadata projects in order to identify what types of metadata exist and how they are used; and
- comparison and analysis of the metadata elements to show how they are related to MARC 21 metadata format elements.

### **Survey and analysis of ongoing metadata projects**

In order to identify what types of metadata exist and how they are used, a survey of the Web and library literature was conducted. Additional information was taken from a questionnaire that was sent to several specialized mailing lists while the author was serving on the "ALCTS" Task Force to study this issue [4]. The Web survey pointed to hundreds of ongoing metadata projects. Many of these are listed in the IFLA Web page [5].

After the initial survey, a number of projects were selected for closer analysis. Since the purpose of this paper is to provide an overview of main trends only, the methodology applied in selecting the projects reflects a tendency to focus on those projects that illustrate what the author perceives to be the main currents in metadata development and application.

Many of the major metadata projects initiated recently have been designed to address specific problems or needs in classifying and cataloging digital resources. Creators of these new metadata standards had specific formats or subjects in mind.

The first such example is the Library of Congress project "BeOnline"[6], which concentrates on business and economics-related materials, especially those which will facilitate business reference in the area of entrepreneurship and small business. This project uses the Library of Congress PCC Core Record, which makes it one of the very few projects to use national or international standards. The purpose of this project is to catalog Web sites and provide access points to these resources. The description of these resources is fuller than Dublin Core but does not follow AACR. All access points (name and subject headings) are established

according to the LC authority control.

In the "BeOnline" project, records are created from data captured from the Web site. All required MARC fields are provided. The data then are converted into MUMS records through the TCEC program. A cataloger edits descriptive portions of records in order to correct obvious errors. A cataloger supplies correct headings and authority records. Keywords are captured or provided by the selector and converted to field 653 (uncontrolled subject term). A subject cataloger provides classification and one or two subject headings. Records are encoding level 7 (minimal) and are labeled "BeOnline metadata record" in 8z (public note) of field 856 (URL field).

Although the project "BeOnline" is a model for applying standards and creating records that can be easily incorporated into any library online catalog, so far only Library of Congress staff are able to create records from data captured from the Web. The Library of Congress ought to consider allowing other institutions, including academic and public libraries, to use the same mechanism in creating records according to this standard.

The "Alexandria Digital Library project" (ADL) [7], focuses on spatial data, and especially on remote-sensing imagery such as aerial photographs and satellite images. The ADL started out using MARC 21, AACR2R, and LCSH metadata, but has added to or departed from these as appropriate for specific material types. For example, about 26 fields that contain technical information for remote-sensing imagery (such as altitude of sensor) have been taken out of the General Note (500) field and each given separate fields of their own, as additions to MARC 21.

Three projects related to art apply different metadata standards. The first project, "Categories for the Description of Works of Art" (CDWA) [8], was developed for art and architecture specialists. It provides a structure for describing works of art and electronic images of them. The standards are based on categories rather than on rules. In contrast, the second project, the "Art, Design, Architecture, and Media Information Gateway" (ADAM) [9] is using established standards and metadata such as cataloging rules based on the AACR2, Olsen's *Guide to Cataloging Networked Resources*, the *Art & Architecture Thesaurus*, the Dewey Decimal System (21st edition), templates, the "Dublin Core Metadata Element Set" and so on.

The third project is "Project Runeberg" (Nordic Literature and Art) [10], which began in December 1992 and has published several hundred titles of Nordic literature and art in full electronic text on the Internet. Project staff scan old books, convert the text to HTML file, and put them up on the Web server at <http://www.lysator.liu.se/runeberg/>. Project Runeberg uses general-purpose Internet and UNIX software. It publishes only literature in Nordic languages, such as Swedish, Danish, Norwegian, Icelandic, and Finnish. It also publishes information about Nordic authors. Project Runeberg is not a library catalog. It does, however, publish the text itself and information about the "collections" included in the project.

The "Australia New Zealand Land Information Council" (ANZLIC) [11], is worthy of mention because of its uniqueness. ANZLIC's strategic plan was to develop and implement a national land and geographic data directory system for Australia and New Zealand. This project designed core metadata elements and core elements, which relate to similar information, have been grouped into categories:

- Dataset: Title, Custodian, Jurisdiction;
- Description: Abstract, Search Word(s), Geographic Extent, Name(s), or Geographic Extent, Polygon(s);
- Dataset currency: Beginning date, Ending date;
- Dataset status: Progress, Maintenance and Update Frequency;

- Access: Stored Data Format, Available Format Type, Access Constraint;
- Data quality: Lineage, Positional Accuracy, Attribute Accuracy, Logical Consistency, Completeness;
- Contact information: Contact Organization, Contact Position, Mail Address 1, Mail Address 2, Suburb or Place or locality Suburb of the Mail Address, State or Locality 2, Country of the Mail Address, Postcode, Telephone, Facsimile, Electronic Mail Address, Metadata Date, Additional Metadata.

The ANZLIC project in a way is similar to the US approach developed by the Federal Geographic Data Committee (FGDC). While ANZLIC has not adopted the this approach precisely, the Australia New Zealand framework is still consistent with the guidelines on Digital Geospatial Metadata produced by the US FGDC and with the Australia New Zealand Standard on Spatial Data Transfer AS/NZS 4270.

The Government Information Locator Service (GILS) [12], is a tool for identification of US government information resources. GILS has developed a complex metadata format as part of this service. GILS is a complex metadata format because the MARC and Z39.50 communities have heavily influenced its design. In the *Application profile for the Government Information Locator Service, Draft version 2* the users will find all of the elements in the GILS Core Element Set. These elements are defined as either being repeatable or not repeatable; the repeatable elements may appear more than once in a single GILS record whereas the not repeatable elements can only appear zero or one times [13]. The Core Element Set goes beyond descriptive, subject, and basic availability information. It includes separate fields for details on the point of contact and the provenance/source of the information, a number of administrative fields, and fields for copyright and other access constraints.

The types of resources described in GILS include databases, hotlines, clearinghouses, and catalogs of publications. GILS records for these resources are not the resources themselves, but are descriptions of the resources which inform users what information is available, where it is located, and how it can be accessed. In cases in which the information resource exists in an electronic form, a direct link to that resource from the GILS record will usually be available [14].

Government Information Locator Service (GILS) Core Elements include:

- Title; Originator; Abstract/Purpose; Distributor; Time Period; Technical Prerequisites;
- Methodology (Instrument) Geographic Name; Methodology (Instrument Platform);
- Discipline; Parameter/Keywords; Project/ Campaign.

GILS also created detail elements (about 36 elements). See their Web page [15].

The Dublin Core project (DC) [16], grew out of a workshop sponsored by OCLC and the National Center for Supercomputing Applications in 1995. The Dublin Core metadata is described as an efficient and simple metadata for electronic articles and digital objects. The Dublin core elements describe a journal article, movie, image etc. The Dublin Core Metadata Element Set represents a simple resource description record. Elements are designed to be used by content creators. The Core contains just 15 metadata elements:

- Subject and keywords: The topic addressed by the work;
- Title: The name of the object;
- Author or creator: The person(s) primarily responsible for the intellectual content of the object;

- Publisher: The agent or agency responsible for making the object available;
- Description: textual description of content;
- Other Agent: The person(s), such as editors and transcribers, who have made other significant intellectual contributions to the work;
- Date: The date of publication;
- Object Type: The genre of the object, such as novel, poem, or dictionary;
- Form: The data representation of the object, such as Postscript file or Windows executable file;
- Identifier: String or number used to uniquely identify the object;
- Relation: Relationship to other objects;
- Source: Objects, either print or electronic, from which this object is derived, if applicable;
- Language: Language of the intellectual content;
- Coverage: The spatial locations and temporal duration's characteristic of the object;
- Rights management: a rights management statement, an identifier that links to a rights management statement, or an identifier that links to a service providing information about rights management for the resource.

Each element is repeatable and optional, and the entire set has been defined as extensible. Each Dublin Core metadata element can also have a sub-type and sub-scheme information. For example, if an existing scheme is being used for subject and keywords, such as the Library of Congress Subject Headings (LCSH), then this information can also be attached to the element name [17]. A complete list of proposed sub-schemes and subtypes can be found at Knight and Hamilton (1997) [18].

Currently, DC is a widely used metadata standard especially in Europe. For the applications and list of projects using DC, see OCLC at <http://purl.org/dc/projects/index.htm>.

One of the most important cooperative projects using the Dublin Core is the "Cooperative Online Resource Catalog" (CORC) project [19]. The project was developed by OCLC and it allows users, librarians, selectors to select the appropriate Web sites and create automatic bibliographic records based on the DC elements. CORC displays records also in MARC formats. Although mapping from DC to MARC for certain fields is not accurate, it is an effective way to create a base record and then upgrade it as appropriate. The CORC project is growing rapidly.

The Federal Geographical Data Committee's *Content Standards for Digital Geospatial Metadata* (FGDC) specifies the information content of metadata for a set of digital geospatial data. The purpose of the standard is to provide a common set of terminology and definitions for concepts related to this metadata [20]. Geospatial metadata allows researchers to locate relevant data sets for use in geographic information systems (GIS). The Geospatial metadata format was developed by the Federal Geographic Data Committee in 1994 and is called the *Content Standard for Digital Geospatial Metadata* (CSDGM) [21].

The FGDC standard is a complex format with over 334 different elements, 119 of which exist only to contain other elements [22]. Explaining these standards is not an easy task. Users of these standards had to be familiar with the organization of these standards and how to apply them. In the *FGDC Metadata Workbook, Version 2* (Adobe PDF - bookmark enabled), the user will find very good directions on how to read and use the standard, the content information of the standard, a description of profiles and user defined metadata extensions, and some metadata examples. In short, "the standard is organized using numbered chapters called 'sections.' Each section begins with the name and definition of the section. These are followed by the component elements of the section. Each section provides the names and definitions of the component

elements, information about the types of values that can be provided for the elements, and information about the elements that are mandatory or repeatable. The standard is designed to describe all possible geospatial data" [23].

Currently, there are over 100 spatial data nodes in the Geospatial Data Clearinghouses. They contain digital geographic data primarily for use in Geographic Information Systems (GIS), image processing systems, and other modeling software. These data collections can be searched through a single interface based on their descriptions, or "metadata" (24). The FGDC standard has also been mapped to many other existing standards such as the NASA Directory Interchange Format (DIF), the Government Information Locator Service (GILS), MARC 21 and the Dublin Core.

The Colorado Digitization Project (CDP) [25], is a collaborative initiative involving Colorado's archives, historical societies, libraries, and museums. The Project is creating a digital library-museum that provides the people of Colorado with access to historical and cultural resources of this state. The Colorado Digitization Project is using the Dublin Core. The CDP has selected seven mandatory elements for describing digitized resources, in addition to desirable metadata elements that an institution might find useful in describing digital resources. There are seven elements which must be included in the description of each digital resource. The elements are shown in Table I.

For additional needed information about a resource, the Dublin Core Working Group has developed a "qualified" Dublin Core, which consists of the element and its qualifiers. The idea is that the basic elements may be further enhanced by use of these qualifiers, with the purpose of informing the user on how to view or interpret the content of the element. These qualifiers are defined as modifiers and schemes.

MARC 21 metadata formats (US Machine Readable Cataloging) [26], in simple terms are standards for the representation and communication of bibliographic and related information in machine-readable form. MARC 21 is an international standard for computer-readable bibliographic records. MARC 21 records contain description derived from AACR2R rules (for North American, British and Australian libraries), "access points" to the record (author(s), title), subject headings assigned from a controlled vocabulary such as the Library of Congress Subject Headings, and a "call number" to group together items in like subject areas. There are many variations on the MARC 21 record, often country-specific. The MARC 21 record involves three elements: record structure, content designation, and the data content of the record.

**Table I** CDP elements for describing digitized resources

Mandatory elements	Optional (desirable)
Title	Contributor
Creator	Publisher
Subject	Relation
Description	Type
Identifier	Source
Date	Language
Format	Coverage Rights

The structure of MARC 21 records is an implementation of national and international standards. Content designation, the codes and conventions established to identify explicitly and characterize

further the data elements within a record and to support the manipulation of those data, is defined in the MARC 21 formats. The content of most data elements is defined by standards outside the formats, for example, Anglo-American Cataloguing Rules, Library of Congress Subject Headings, National Library of Medicine Classification.

An alternative to using the very complex MARC 21 format to catalog electronic resources was provided by the Program for Cooperative Cataloging (PCC). The "Core Level" standard for cataloging was and is intended to be an intermediate step between minimal and full levels. "Core Level" cataloging contains a complement of data elements necessary for the creation of useful records. It is smaller than that for full level cataloging but more extensive than that for minimal level cataloging. For books, core level cataloging essentially calls for fewer notes, fewer subject headings, and fewer added entries[27]. The core provides some flexibility in adding additional elements. This depends on the cataloger's judgment.

The basic value of the PCC core record is that it provides an acceptable quality of records, which could be shared by libraries and non-library agencies without revision. Since the rules and standards of the PCC are now simplified, it will be possible to create a great many quality records with relative ease by applying the PCC. If a full bibliographic record is required by a library to meet certain needs, the library can upgrade the PCC generated record to a full bibliographic record. Most access points in the PCC core records are in authorized forms and supported by authority work. The PCC core records are generally of such high quality that they require little additional work before they can be incorporated into the OPAC.

The PCC has developed a separate Core standard for books, graphic materials, moving image materials, music and non-music sound recordings, printed music and music manuscripts. A Core record has also been developed for non-Roman script materials such as Japanese, Arabic, Chinese, Korean, Persian, Hebrew, Yiddish, and for serials.

PCC Core Level records can be identified by CC: 9410, the code pec in the 042 field (042 fields show in MUL displays), and the encoding level 4. (See example of the PCC Core in the Appendix.) The PCC Core is currently used to describe electronic resources and Web sites in the LC project "BeOnline". In this project, as described earlier in this paper, LC is creating machine driven cataloging records based on the PCC core.

This brief survey testifies to the variety of metadata standards. Information about additional metadata standards which are not discussed in this paper is available upon request from the author. Some of the standards reviewed in this survey were created to meet specific discipline needs, while others addressed issues of geographic regions or languages. Some standards preferred to use traditional and existing standards and methods of providing access to Internet resources and Web sites while others created their metadata based on categories.

### **Comparison and analysis of metadata elements to each other and an attempt to show how they relate to the MARC 21 metadata format elements**

Four metadata projects were selected in order to compare them to MARC 21 metadata fields to determine if the new electronic formats, such as Web pages, electronic journals, etc., require the development of a new set of standards. These projects are the "Government Information Locator Service" (GILS), "Dublin Core" (DC), "Standards for Digital Geospatial Metadata" (FGDC), and the "Colorado Digital Library Project" (CDP). These projects have been selected because they represent a variety of ways of creating metadata and because they were developed by different types of institutions, ranging from the government to the academe.

In order to analyze the information, the author made use of existing crosswalks[28], and

designed a crosswalk to link MARC 21 to GILS, FGDC, DC, and CDP (Crosswalk of MARC21, GILS, FGDC, DC and CDP is available upon request from the author). This crosswalk allowed mapping the metadata elements from GILS, FGDC, DC, and CDP to MARC 21 metadata and to compare the properties of the various fields used in the metadata standards. Application of the crosswalk analysis produced results described below.

## **Description**

In examining the five metadata standards, one is struck by the contrast between the relative complexity of some (MARC 21, GILS and FGDC) and overall simplicity of others (DC and CP). It appears that both the Dublin Core and its derivative Colorado Digital Project aim at producing what amounts to a short record form for digital formats. CDP is based on DC and simplified or reduced the DC to the mandatory fields and optional fields. Both standards are based on the familiar features of the MARC 21 format, although there are some differences in labeling and data organization. The GILS and the FGDC, on the other hand, are designed to address the needs of specialized digital formats that convey data through methods unknown to the original designers of MARC 21. These two standards appear to go beyond the MARC 21 and create special fields for specialized data.

## **Terminology**

A lack of common terminology currently exists among the different metadata standards. In MARC 21, metadata are identified using the term <tag>, <indicator> and <subfield code>, whereas Dublin Core identifies its metadata using the term <label>. In another example, CDP uses the terms <creator/generator> where MARC 21 uses <author>, and GILS uses "Originator."

MARC 21 qualifies each element as repeatable or non-repeatable. CDP standards indicate whether or not an element is mandatory or optional. FGDC incorporates both of these attributes into a single property by indicating a lower and upper bound on the number of times an element may occur.

These variations in terminology made the process of mapping these metadata elements and creating the crosswalk very difficult. It is not clear why these innovations were introduced by the creators of the new standards. It seems that the information community would be better served by an overall effort to maintain uniform terminology among all standard producers, whenever it is reasonable and practicable.

## **Derivative**

In mapping the metadata elements, it was surprising to see that all the elements were corresponding to the MARC 21 format equivalents to some extent. All of them were covering the same elements in MARC 21 (such as author, title, etc.) regardless of whether the new metadata standard was simple or complex or whether the terminology was changed.

In DC and GILS the "Contributor" field is used to designate an entity responsible for making a contribution to the content of the resource. This field is very general. However, in the MARC 21, "Contributor" it is used in a very specialized way. For example:

- 700 1#\$a Added entry - Personnel Name with \$e=collaborator;
- 710 2#\$a Added entry - Corporate name with 8e=collaborator;



- 711 2#\$a Added entry - Conference Name with e=collaborator;
- 720 ##\$e Added entry - uncontrolled Name/Relator term.

## **Extendable**

The Dublin Core is currently proposing to add another set of "qualifiers" to make the DC more extendable to meet certain needs. This feature of development will add more complexity to the DC and will make it look like MARC 21 under a different name.

## **Different uses of fields**

In CDP, the 6xx field is used for describing the topical subject headings associated to people, places, events, or other contextual information. In MARC 21, the 6xx is detailed to reflect specific information such as a personal name subject heading, a corporate name subject heading, a topical subject headings and so on.

## **From general to specific**

The general note field was labeled differently in the four standards. In GILS, it is used for supplemental information, the DC used it for coverage, the FGDC used it for purpose, and CDP used it for description. The four projects used this field for general notes that could include abstracts, or other description not included in other fields. On the other hand, MARC 21 used 5xx for a specific note and developed in the field 5xx all kinds of notes, which are not represented by any other field.

In CDP, the MARC 21 fields 856, 020, and 022 are used for the "identifier", which is to be used to identify the digital resources. The identifier field includes the International Standard Book Number, the International Standard Serial Numbers (ISSN), and the accession number. The CDP chose to merge the unique identifiers of three separate MARC 21 fields into a single field, which is repeatable.

The MARC 21 standard made a great effort to distinguish between the personal name of the author and a conference name, which is used as a corporate author. This information is contained in fields 100, 110 and field 111 respectively, using added entry for both, whenever it was appropriate. This method of separating the two identifiers makes information retrieval easier. CDP does away with these distinctions and merges the information into two fields: "creator and contributor".

In CDP, the Contributor field is optional but means the same as in the DC. In CDP, the lxx is used to represent the "Creator" field, which is a person or organization, primarily responsible for creating the text. This field is mandatory and repeatable. In MARC 21, this field is specialized for certain kind of authoritative responsibility. For example, 100 10 field is for the author's personal name, 110 field is for corporate name and so on.

An element of confusion is introduced by the authors of a standard when the contents of a single MARC 21 field are used for many purposes. For example, CDP maps its "Identifier" field and its "Format" field to the 856 field in MARC 21. In this case it is very difficult to determine which information goes in what field.

Access constraints have been mapped differently by the creators of GILS, CD, FGDC, and CDP. CDP, for example, mapped this field to the 506 MARC 21 field, while GILS, CD, and FGDC mapped this field to MARC 21 540 field. These two fields in MARC 21 seem to have the same meaning and it is difficult to know when one is to use them (506 and 540 fields). The

terminology used in these fields in particular is very confusing. For example, CDP uses "right management", GILS and FGDC use "use constraints" and DC uses "right".

In examining the crosswalk one notices that MARC 21 fields 006-019, 025-033, 046-099, 210-243, 254-257, 261-265, 300-307, 521-533, 584-599, 730-785, and 790-852 have no corresponding fields in GILS, DC, FGDC or CDP. This indicates the sophistication and the level of detail offered by MARC 21 as compared to the other standards. Other evidence suggests that the other standards merged some of the fields together, creating more general fields.

On the other hand, evidence derived from the crosswalk chart suggests that FGDC employs a number of fields to contain detailed information required by the discipline it serves. For example, the field 342, with all the sub-fields, is used to contain information about geospatial references. Similar information is included in field 352 with its sub-fields, and the 514 field and its sub-fields. In comparing the GILS to FGDC, one discovers that they both use note fields to provide special information.

A summary of the analysis can be encompassed in the following general statements:

- This study attempts to analyze some features of the four commonly used metadata sets by comparing them with MARC 21. Comparisons of the different ways in which each set of standards deals with information provided above, suggest the potential difficulties a user will have in identifying his information sources. At times it was difficult to establish exact relationships among the different metadata standards analyzed here. Given that there are already many more metadata standards and more are generated every day, the results may do little to improve access.
- The obvious difficulty is that a single resource will be described by two or more sets of unique metadata attributes, which will lead to the creation of multiple and distinctly different, if not altogether contradictory records. Searching for resources described bibliographically in a variety of ways will cause significant difficulties for the end user.
- More potential difficulties arise when the standards are revised or extended. As new types of resources are defined, or new applications developed, it will become necessary to extend the metadata standards. As these changes will invariably occur, it will be difficult to constantly update and revise the crosswalk. It will also be difficult to transform the metadata from one set to another.
- Another implication of different standards stems from the role of the vendor or the designer of the system. Every time the metadata standards are extended to represent local information there will be a need to work with the vendor to accommodate these changes.

## **The need for metadata**

Why do libraries need metadata in the era of digital electronic resources? To address this question, it is necessary to keep in mind that we are still dealing with information that is created for the same purpose as the book and the journal, and therefore, subject to the same classification and description laws as the rest. Libraries use traditional tools to organize information, which facilitates access to print and non-print materials. AACR2, LC classification, Dewey, LSCH, and other tools have been used and continue to be used to describe and classify information. Libraries are likely to continue to develop and improve these old standards as they adapt to the demands of the publishing and information industries.

Since electronic information behaves differently than the print media, it requires special treatment by those describing and classifying it and this is where metadata comes in.

Experience with using electronic resources suggests that to facilitate the creation of a basic catalog record for an electronic resource, it is useful to have the metadata embedded in the electronic text itself. Once this catalog record is created, it can be enhanced or edited or used as it is by any library that chooses to add it to its catalog. These records could then be added to the library online catalog or can stand alone, depending on individual library's needs.

The major uses of metadata are to:

- organize and maintain an organization's investment in data;
- provide information to data catalogs;
- provide information to aid data transfer; assist effective discovery and retrieval of information and allow users to determine if the data they find is what they want;
- control restricted-access information and prevent some users (like children) from accessing data;
- provide common agreement on what elements to use or what their content should be;
- give information that affects the use of data, such as legal conditions on use, its size, or age;
- give the history of data, such as the original source of the data and any subsequent transformations;
- give information about the owner or the creator of the text; and
- give relationships with other resources such as linkages to previous and subsequent versions.

Other uses of metadata have been suggested. Milstead and Feldman provided other usage of metadata:

To make sure that all the materials about the same subject were found together either on the shelf or in an online database; To single out important concepts from those which are merely incidental to the work; To ensure that the same information was found for each work, and that it was put in the same place, so that someone searching for works by an author named Fields would not find them mixed with agricultural tracts on fertilizing wheat fields; To provide proper use of indexing vocabularies and field structures, both in searching and in cataloging, increases precision and minimizes the chance of false drops [29].

Metadata standards play an important role in supporting the use of electronic resources and services. In order to maximize its use and assure consistency, it needs universally accepted standards, which would govern its creation.

## **Conclusion**

Metadata is simply a method used to describe an electronic document (data) and it can be done at the point of creating the document or afterwards. Experiments with creating new metadata standards provided simpler short-form standards that could be used by the creators of electronic data sets and documents. This allows for the creation of metadata at the time of the creation of the object itself, either by or under the auspices of its creator. The more traditional method of adding metadata at some later point, by library cataloging staff or commercial services remains an option.

Dublin Core, GILS, FGDC, and CDP are among the best known metadata sets that have been created to describe electronic resources, but they are only the tip of the iceberg. Many other metadata projects are being designed to address this need, and descriptions of some of the more

important ones have been provided here. Some of standards are being created for specific disciplines or non-text electronic files, while others are designed for more universal applications. This preponderance of metadata sets offers an unprecedented flexibility in describing electronic data and should lead to improved access to the documents themselves.

Paradoxically, this diversity in classification also contributes to the growing problem in accessing various data sources. The problem arises from diversity itself. It appears that many creators of new metadata standards are keeping an eye on the MARC 21 standard, but pick and choose among its elements and rename its tags willy-nilly. The resulting loss of a single coherent standard adds to the problems of accessing electronic resources.

The Library of Congress has traditionally played an important role in creating standards for cataloguing. Libraries became familiar with these standards and have been very successful in organizing their collections according to these standards. The user has also benefited greatly from this standardization, which made information retrieval relatively easy. Claims that the user doesn't understand the catalog or doesn't care much about what standards are used have little bearing on the overall benefit derived from the single system used to organize knowledge.

Libraries are currently using the LC standards to describe their books, periodicals, microforms, and other formats and it seems that the same system can be used to describe electronic resources. The LC standards are universally accepted and perhaps some thought should be given to a return to the basics, as we attempt to describe electronic resources.

The Library of Congress introduced a new initiative to create a standard for electronic resource cataloging. The PCC Core is attempting to address the special needs of electronic formats, which are not met by the currently existing standards. Libraries have an opportunity to take a closer look at the "Core" and decide whether it meets their cataloging needs.

The benefits of using a standard which is based on MARC 21 and which conforms to its structure are many. There is potential to eliminate the problems and the confusion of terminology, the expenses in using several standards and then seeking technological solutions to the problems this creates. There is every expectation that this new standard will greatly increase the effectiveness of retrieving the needed information. If this project succeeds, its element of description will form the future metadata standard.

As the library profession evolves in response to changes brought about by the developments in information technology, we librarians must not forget our roots, and more importantly the established principles of information classification which have developed into standards that govern the work of this profession until today, and for the foreseeable future.

## ***Notes***

- 1 *A User Guide for Simple Dublin Core*, Draft version 5.1 Revised 31 July 1998, *Online*, Available: [http://purl.org/DC/documents/working\\_drafts/wd-guide-current.htm](http://purl.org/DC/documents/working_drafts/wd-guide-current.htm)
- 2 Computer Wire: Data Warehousing Tools Bulletin: Briefing Paper: What is Metadata. Available: [http://www.computerwire.com/bulletinsuk/212e\\_1a6.htm](http://www.computerwire.com/bulletinsuk/212e_1a6.htm)
- 3 Jessica Milstead and Susan Feldman, "Metadata: Cataloging by Any Other Name", *Online*, January 1999, Available: <http://www.onlineinc.com/onlinemag/OL1999/milstead1.html>
- 4 Judith R, Ahronheim and Magda El-Sherbini, "Magda and Judy Web page", Available: <http://www-personal.umich.edu/~jaheim/alcts/email.htm>
- 5 IFLA Web page, Available: [www.ifla.org](http://www.ifla.org)
- 6 Business and Economic Resources Online, Available: <http://www.loc.gov/rr/business/beonline/meta.html>
- 7 Alexandria Project (Alexandria), Available: <http://www.alexandria.ucsb.edu/~lhill/>
- 8 *Categories for the Description of Works of Art (CDWA)*, Available: [http://www.gii.getty.edu/index/\\_cdwa.html](http://www.gii.getty.edu/index/_cdwa.html)
- 9 Art, Design, Architecture, and Media Information Gateway (ADAM), Available:

- <http://adam.ac.uk/adam/metadata.html>
- 10 Project Runeberg (Nordic Literature and Art), Available: <http://linnea.helsinki.fi/meta/nmfinal.htm>
  - 11 The Australia New Zealand Land Information Council (ANZLIC), *Core Metadata Elements for Land and Geographic Directories in Australia and New Zealand*, Available; [http://www.anzlic.org.au/\\_metaelem.htm](http://www.anzlic.org.au/_metaelem.htm)
  - 12 Government Information Locator Service (GILS), Available: <http://www.usgs.gov/gils/>
  - 13 *Application profile for the Government Information Locator Service*, Draft version 2, Available: [http://www.usgs.gov/gils/prof\\_v2.html](http://www.usgs.gov/gils/prof_v2.html)
  - 14 *What is the Government Information Locator Service (GILS)?*, Available: <http://www.epa.gov/earth100/docs/whatis.html>
  - 15 Government Information Locator Service (GILS), Available: <http://www.usps.gov/GILS/usps0012.htm>
  - 16 Dublin Core (DC), Available: [http://purl.org/metadata/dublin\\_core](http://purl.org/metadata/dublin_core)
  - 17 Renato Iannella and Andrew Waugh, *Metadata: Enabling the Internet*, Available: <http://www.dstc.edu.au/RDU/reports/CAUSE97/>
  - 18 Jon Knight and Martin Hamilton, Dublin Core Qualifiers, Available: [http://www.roads.lut.ac.uk/\\_Metadata/DC-SubElements.html](http://www.roads.lut.ac.uk/_Metadata/DC-SubElements.html), 1997.
  - 19 Cooperative Online Resource Catalog (CORC), Available: <http://www.oclc.org/oclc/promo/10520corc/index.htm>
  - 20 The Federal Geographical Data Committee's Content Standards for Digital Geospatial Metadata (FGDC), Available: <http://www.fgdc.gov/>
  - 21 Ed Summers, "Show me the Metadata!", *Online*, Available: <http://libstaff.lib.odu.edu/~esummers/meta/meta.html>
  - 22 *Tools for Creation of Formal Metadata*, Frequently asked questions on FGDC metadata, Available: <http://geology.usgs.gov/tools/metadata/tools/doc/faq.html#standard>
  - 23 The Federal Geographical Data Committee (FGDC), *Metadata Workbook*, Available: <http://www.fgdc.gov/metadata/metadata.html>
  - 24 Geospatial Data Clearinghouse, Available: <http://130.11.52.178/>
  - 25 Colorado Digitization Project (CDP), Available: <http://coloradodigital.coalliance.org/>
  - 26 *USMARC Format for Bibliographic Data*, Prepared by Network Development and MARC, 1994, Standards Office, Cataloging Distribution Service, Library of Congress, Washington, DC.
  - 27 Program for Cooperative Cataloging (PCC) Core Record, Available: <http://lcweb.loc.gov/catdir/cpso/corelev.html>
  - 28 Margaret St. Pierre and William P. LaPlant, "Issues in cross walking Content Metadata Standards", *Online*, Available: <http://www.niso.org/crswalk.htm>
  - 29 Jessica Milstead and Susan Feldman, "Metadata: Cataloging by Any Other Name", *Online*, January 1999, Available: <http://www.onlineinc.com/onlinemag/OL1999/milstead1.html>

## ***Appendix. Summary of the PCC Core Record***

- 020, (ISBN, ISSN) Mandatory if present on item.
- 040 (Cataloging Source) Mandatory.
- 042 (Authentication Code) Mandatory.
- 050,082,086, etc. (LCC, Dewey, Government Document, etc. classification numbers) Mandatory. Assign at least one classification number from an established classification system recognized by USMARC. Assignment in some cases will constitute solely classification while in others it will consist of a full call number.
- 1XX (Main Entry) Mandatory if applicable.
- 240 (Uniform title) Mandatory if applicable and known or readily inferred from material being cataloged.
- 245-300 (Title Statement; Edition Statement; Publication, Distribution, etc.; and Physical Description) Mandatory. Describe fully, using all data elements appropriate to the item described.
- 490 (Series Statement) Mandatory if present.
- 440/8XX (Series Added Entry) Optional to make added entry for a series. If a series is traced, it must be in an authoritative form supported by a national-level authority record. Other aspects of series treatment are local in nature.
- 5XX (Notes)

Optional, except the following. The basis for the recommendation is that only those notes that support identification of an item need be included. These criteria will vary from one form of material to another. With respect to justification of added entries, that function may be provided alternatively in some cases through other data elements,

e.g., the tagging itself or use of relators.

- 500 (General Note) Required for source of title if not from title page.
- 505 (Contents Note) Required for multipart items with separate titles.
- 533 (Reproduction note) Required for reproductions.
- 6XX (Subject Access Fields) Required if appropriate. Assign 1 or 2 subject headings at the appropriate level of specificity from an established thesaurus or subject heading system recognized by USMARC. More than 2 may be assigned.

- 7XX (Added Entries)

Required if applicable. Assign 1) a complement of added entries that covers at least the primary relationships associated with a work (e.g., joint authors); 2) added entries to bring out title access information judged to be important. For both 1) and 2) above, determination of primary relationships and of the relative importance of title access information is intended to reflect either individual cataloger's judgment or the institutional policy of the participant.

- 8XX See 440/8XX above.